

electronic publishing in science

Proceedings of the Joint ICSU Press/UNESCO Expert Conference
February 1996

Conference Chairman: Roger Elliott
Steering Committee Chairman: Kai-Inge Hillerud
Programme Committee Chairman: Dennis Shaw
Local Committee Joint Chairmen:
André Authier and Howard Moore

Edited by Dennis Shaw
and Howard Moore



Options for the future

Joshua Lederberg

*Sackler Foundation Scholar and
President Emeritus, Rockefeller University, New York, USA*

First I wish to review what the scientific literature is all about and, in all my remarks, I'm making a very explicit distinction – as my colleague [de Kemp] did before – that our discussion concerns principally the primary scientific literature, the original reporting of scientific data and theory, formulation and assertion of claims with respect to priority, and the like. I think a very different set of rules applies to that literature than to the dissemination of monographs, textbooks, novels, biographies and so on.

It is a special characteristic of that primary literature that its authors, on the whole, are totally uninterested in royalties, which indeed have generally not been available to them. Their gain from publication is recognition by their colleagues and the dissemination of knowledge in the spirit of science.

Primary literature constitutes, as it stands on the shelves of libraries and in other formats, an inerasable public archive. I think it is of the utmost importance that we not take fluidity too lightly, that there be a point of commitment when the author says: 'These were my words; this is what I said we have done in the laboratory; these are the claims that I am making and they should not be tampered with at any further stage.' Obviously they may be modified, there can be links to corrections addenda, further distinctions, but it should always be possible to reconstruct what the author is to be held accountable for.

The registration of claims is a very important point – what drives science is the possibility of making a novel discovery. It's not a novel discovery if somebody else made it first, and we get very scant credit for even being

hard on the heels of others, who had managed to get there a few minutes before. There are many implications of the allocation of scores for achieving success in that regard, but it's built into the structure of science as innovation that there be a system of registration of claims for what is new, what is different, what is distinctive, that was 'my contribution' to the growing corpus of scientific knowledge and understanding.

A scientific publication is a grave act to be undertaken with the utmost seriousness; it's an inscription under oath. To lie in a publication is *de facto* perjury, and, when that is discovered, there are consequences no less serious than perjury in court and so it should be. We shouldn't have to worry about whether an assertion of data from an experiment or other claims was made with other than the utmost seriousness on the part of the author. Otherwise, we would be eternally confused whether the matters presented deserve our attention. And the literature is a historic repository where the record of our scientific culture can be refreshed and re-examined for the purposes of history, for the purposes of rediscovering ancient matters, whose significance was not fully understood before, the establishment of links between different disciplines, and so forth.

So you see, I take the literature very seriously. To me it's holy writ and I want to be sure that, in whatever format it is distributed, it will be accessible, its veracity can be attested by its being observable by everyone at will, and it should be an achievement that cannot be altered. But it's also a dynamic place, as it should be, and effectively is an open forum. It could be more open than it is. We don't have enough dialectic in our current modes of publication after the fact. There has to be almost a

federal case before you can have a comment published on a prior article, and the electronic media will help to lower the threshold in that regard. It's a 'rumen', which is a place for fermentation, for digestion, for re-examination of the given truths over a period of time, and it certainly has a dynamic quality taken in its totality, even if every brick of the edifice has been put firmly into place in a severely qualified process.

And then, of course, a very important social function – perhaps the most important function that the print journals now have and would be difficult replacing with the electronic media – this is the dignity of the work, having a physical representation of clear type on durable paper. That goes along with the attributes that I mentioned earlier, being a definitive act of publication in its original form, and so forth. That dignity is attested by the imprimatur of the editors. There's no reason in the world why that intrinsic function cannot be transferred to the electronic forum. This question has consequences for the accumulation of prestige, for competition for tenure, for grants, for attraction for students – all of the things that are involved in the social system of science and that make science work. We have to be very careful in guiding the evolution of the electronic journal, to be sure that the positive values of our current system are preserved while maintaining lively access to information – which is the primary asserted advantage of this new form.

We have a community of actors concerned with scientific publication. They have sometimes convergent, sometimes divergent interests. The authors above all want to make their work known to others. That is their superordinate goal in publication. In order to achieve it, they are certainly interested in having their work available at the lowest price and cost to their readership and, with varying success, this has been internalized in the practice of page charges, which are negative royalties. These are payments that authors are willing to make in order to help lubricate the system, in order to help have their work made available to the community and to others. They're concerned about speed, for a wide variety of reasons: to be sure there are no artifacts in the competition for priority or just to accelerate the process of dissemination so that others can most promptly take advantage of and build on the work that has been published. One is really quite irked at having to wait six, eight, ten months from the time of submission of an arti-

cle – those are typical intervals of delay – and we all believe that whatever we have just done and just submitted is of the utmost importance and that the world is holding its breath waiting for it. Sometimes it is! We look for places that have a reputation for reliability, so that people don't have to waste an undue amount of time just in a first order assessment of whether the statements that are made are to be taken at face value.

As part of that process, the author is as interested in easy retrievability as is the reader. If the reader's searching for material is facilitated by the system, the author knows there will be more readers who will be able to make use of a given contribution. And, yes, one of the things that authors can often profit from is good advice from good editors about their modes of expression. I will never forget the first manuscript that I wrote that was seriously edited. It was covered with blue pencil marks, and I wish I still had a copy of that markup because there was a lesson in every line of it.

Libraries and publishers are increasingly at odds with one another. Pricing policies are leading to a black hole – especially with journals with an already limited subscription – the subscription base goes down, the subscription fees go up and there's a necessary reaction of further retrenchment. The *reductio ad absurdum* will be a single subscription that will cost a million dollars and we will rely on interlibrary loan subject to rules of access for fair use to get copies of it. We're approaching that situation with some journals and of course, it's pure silliness and why print it at all in this case? In other words why not make the transition to the electronic medium forthwith, when we're heading that way anyway?

The reader wants access but not to be flooded with information and, of course, it's hard not to be inundated with material, even just that which is essential and important to one's scientific existence, not to mention everything else that comes to one's attention. We want quality assurance so we look for assistance, we look for filters, we look for peer review as a method of assisting us in navigating through the blizzard of available information that will only worsen. It will worsen anyhow with the accelerating pace of scientific activity, and the ease of deposit on the Internet will aggravate it further. So readers are going to need that kind of filtering assistance more than ever, and that's the big argument for the

peer review imprimatur in some form or another, roughly analogous to what we have with print journals. If we don't get it by a formal method, we will simply have a universe of readers who will have no time to read and they will wait for the equivalent of the book reviews and the commentaries of others – a sort of *ex post facto* review, so you will eventually get around to being told by your colleagues that so-and-so was the article you should have read.

The producers, the publishers and the editors, of course, have had their very important part to play, and their role in quality improvement and quality control has been indispensable. They have established the mechanisms for editorial review on which we rely for the kind of filtration that I have described.

And then – not sufficiently mentioned – there are other people who have an interest in the substantial outcome of scientific work, the sponsors and the beneficiaries. These are the people who pay for the scientific work and the public which is the ultimate beneficiary of that activity. They pay the bill and have every right to expect that we scientists have organized systems which will operate with reasonable smoothness and efficiency: that we don't waste time doing work that's already been done, that we are properly informed about our colleagues' work. That is the assertion that we have made in applying for our research grants and which ensures their worth to the community.

Now there are other functions of literature for the future which are based, on but go beyond, the concept of the intelligent agent for retrieval. I foresee the point at which it will be both possible and necessary to have intelligent agents examining the content of literature and assisting us in drawing of inference, in finding relationships between facts, in truth maintenance – in other words in consistency-checking among the data that are present in the literature – the kinds of things that we exercise by cerebral management with some assistance today. The scientific world has become so complicated – certainly in the biological area – that no single mind can encompass it all properly and we will need conceptual models that involve some computer assistance in dealing with it. Without direct access to the literature on the part of these intelligent agents, the process is quite hopeless. I spent several years in a programme at Stanford with Ed

Feigenbaum, Bruce Buchanan and Carl Djerassi on artificial intelligence in chemistry – the Dendral programme – and by far the most arduous part of that was the knowledge transfer from the factual information from the mind of the chemist – mainly Carl Djerassi – into the set of rules for the computer programs, and that simply is not workable as a practical mode of developing computer systems in other disciplines. We said then and still have to stress that, until the day comes that we have intelligent agents capable of abstracting that knowledge directly from its existing repository, namely from literature, we are going to be in a *cul de sac* for really serious further development in that field. Now all of my remarks so far have come from the journal as the starting point – the existing framework of scientific communication. I'm going to take a slightly different tack at this point.

Recall that the Scientific Journal began as letters of correspondents. They were then aggregated in order to make it a bit more convenient for a given scientist to communicate with his colleagues in those days and so the journals were then founded. We are seeing a re-enactment of that process today with electronic mail, bulletin boards, discussion groups and lists and they're here to stay. I'm not making any exhortations or predictions about the future. I'm talking about current circumstance.

Scientists and other scholars are on a rising wave of exercise of that intercommunication capability. Most of it is rather informal. I'm a little worried that, if this process goes on without some discipline, that it will be all too fluid and some of the virtues of accountability and of seriousness, of freedom from pollution and from exploitation by individuals who have nothing better to do than run their mucky fingers over the typewriters all day long and so forth, will create such a blizzard of junk that we will not be able to find our way through it. So, if we had never heard of the Scientific Journal in its print form, and were just watching the manifestation of this communication as it is operating today on the net, I think we would very quickly come to the conclusion we had better invent something like a peer review journal in order to provide some modicum of order and of discipline in that medium.

My recommendation is that we, as professional scientists, really have got to get together and work out a code of conduct about what we regard as responsible scientific communication through the electronic media. What are the appropriate modalities? What are the appropriate forms of submission, examination, tagging, labelling? In asking such questions I think we will discover something very close to what has evolved in the print journal system. The professional societies have a uniquely important role to play in this process.

Now what are some of the foreseeable consequences? I really have nothing to ask of the print publishers or of the 'for profit' electronic purveyors. Unless they are very selective – and they sometimes will be – about their value added, they will fall of their own weight as scientists become empowered to manage their own communications without the benefit of intermediaries. Yes, I'm echoing Paul Ginsparg. If publishers insist, as some of them have been, on defying claims of fair use to mitigate access to material through copyright, they will only accelerate the resentment of the creator-authors, whose primary purpose is dissemination of knowledge and easy access by others to their intellectual output. Some journals, print and otherwise, will be so invaluable, so difficult to replicate, have such a fine customer base, that they will still be wonderful bargains and they will thrive: *Science*, *Nature* and a dozen others that could be named are outstanding examples. But the publishers no longer have a captive audience that has no place else to go. It used to be that subscriptions were automatically taken for every journal by every library, regardless of price and, of course, we know that world has changed irrevocably at this time.

These issues will just work out in future of their own accord as scientists manage their affairs in their own best interest, but one certain source of conflict, regrettably bound to be exacerbated, is the disposition of historic copyright – the many decades of copyrighted scientific material which is now the property of corporate publishers, to whom copyrights had been transmitted, as part of the routine contract that few of us ever stopped to look at, in order to get our papers published.

Obviously, much will depend on the definition and applicability of fair use. We were told that moral arguments played a considerable role in the judicial process

in Michigan, so my second recommendation is for an ICSU/UNESCO-Industry Conference to try to map out some standards of definition that may be more convergently acceptable to both the academic and the commercial communities involved in this process and try to minimize some of the rancour which is otherwise bound to be exacerbated. Not to mention government funders of the underlying research.

So I foresee the evolution of a mixed system in which there will be communications over a broad range of formality, from private to public. I fear that we will see a few invisible colleges that will be closed enclaves. I hope that that can be discouraged as contrary to the true spirit of science. Obviously, there's room for very specific and short-lived collaborations but we none of us want to see an exclusive model develop, and that could be an issue to be taken up under the codes of conduct that I had indicated earlier. There will be people who post their own data and own images from time to time, without authentication through peer review, and these same individuals will try to acquire a little more dignity at other times in presenting their more formal results for the peer review process.

Scientists are moving rather quickly in this field and I think we should, sooner rather than later, try to get some explicit agreement on these codes of conduct. One of the issues to be addressed is the dignity to be given to the stamp of approval. The process seems fairly obvious and will resemble very much our current one – postings that are submitted through this process deserve a mandate to be acknowledged and cited by others. They should get full credit in the various gate-keeper systems, and we haven't turned the corner on that just yet. At this point, few of my academic colleagues would, by preference, submit an article to an all-electronic journal as opposed to the print journal because they are not yet assured that that has the printed journal's prestige when it comes to academic review, when it comes to visibility in other regards and when it comes to grants. It shouldn't be long before that's turned around but I think it would help to have some formal endorsement of that process.

Electronic materials need to be archived – the formidable technical problems of retrieving data from 'dead media' have been eloquently put forth at this confer-

ence. Here again, I can really see no other reliable place where that commitment is likely to be enforced over any substantial period of time but the professional societies themselves. There may be functional agents acting on the societies but I think the moral and contractual commitment for that preservation and perpetuity will have to come from within the scientific community itself.

There will be a second tier of publication and perhaps we ought to acknowledge what somebody – I think Dr Pullinger – remarked that *Nature* sees a lot of articles and they all get published somewhere else anyhow, that nothing is ever really rejected. Well let's admit that. Let's at least examine one possibility (which I haven't really thought through). The result of a review process might be: 'Well, by all means this belongs in our premier journals' or 'This is so obviously faulty that it can't appear anywhere. It is full of slander, it is full of lies. We wouldn't dare touch it with a ten-foot pole.' But the majority of submissions would probably fall somewhere in between and I would say: why not accept them but not put them in the premier site? We will not call it Grade B but everyone will know that is what it means; if you did not make it in the first one, yes, you can have more or less automatic accessibility to the second one. Who will ever read it, who will ever want to look at it? That is another

story, but at least it will be part of the available public record with the advantage of open accessibility, and we will have, I think, saved a lot of churning in going to three or four other journals in order to achieve that result.

One of the aspects of electronic publication is we no longer have any excuse for rationing input. It matters little if the journal is five, ten or twenty thousand pages and particularly if, as I hope will be understood, we use page charges as the primary medium of financing that kind of a system. One of the other benefits of the page charge is that, yes, the potential polluter of excessive input will at least have to pay for the cost of what's coming in.

Well, there are other formats for deposit but I think they would follow fairly naturally from the expectation that there will be a continuum, so my final remark really is that this system is evolving, whatever we say or do or resolve at this Conference. It is going to march ahead anyhow. None of us can be a King Canute saying this tide can't come in and the real issue is: can we channel this technologically driven but, I think at this point, quite inexorable process in ways that will be of the utmost benefit to all the members of our community?